DR. ELIZABETH CLAIRE GOODE (Orcid ID : 0000-0002-8425-1530)
DR. SUN-GOU JI (Orcid ID : 0000-0001-8652-6318)

Article type : Original

Factors Associated With Outcomes of Patients With Primary Sclerosing Cholangitis and Development and Validation of a Risk Scoring System

Authors

Goode EC (ecg44@cam.ac.uk)[1,2,3,4,5], Clark AB (Allan.clark@uea.ac.uk)[4], Mells GM (gfm26@medschl.cam.ac.uk)[2], Srivastava B (dr.bs79@gmail.com)[2], Spiess K (ks460@medschl.cam.ac.uk)[2], Gelson WTH (William.gelson@addenbrookes.nhs.uk)[5], Trivedi PJ (P.J.Trivedi@bham.ac.uk)[6,7,8], Lynch KD (K Kate.Lynch@ouh.nhs.uk)[10], Castren E (edit.castren@gmail.com)[1], Vesterhus MN (mette.vesterhus@hotmail.com)[11,12], Karlsen TH (t.h.karlsen@medisin.uio.no)[11,12], Ji SG (Ji.sungou@gmail.com)[3], Anderson CA (ca3@sanger.ac.uk)[3], Thorburn D (douglas.thorburn@nhs.net)[13], Hudson M (Mark.Hudson@nuth.nhs.uk)[14], Heneghan M (michael.heneghan@nhs.net)[15], Aldersley MA (mark.aldersley@leedsth.nhs.uk)[16], Bathgate A (andy.bathgate@luht.scot.nhs.uk)[17], Sandford RN (rns13@cam.ac.uk)[2], Alexander GJ (gja1000@doctors.org.uk)[5,13], Chapman RW (Roger.Chapman@ndm.ox.ac.uk)[10], Walmsley M (martine@pscsupport.org.uk)[18], UK-PSC Consortium, Hirschfield GM (G.Hirschfield@bham.ac.uk)*[6,7,8,9], Rushbrook SM (Simon.rushbrook@nnuh.nhs.uk)*[1,4].

*Jointly supervised study, experimental design/conduct, analysis and reporting

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/hep.30479

Affiliations

[1]Norfolk and Norwich University Hospital, Norwich, UK; [2]Academic Department of Medical Genetics, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK; [3]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; [4]Norwich Medical School, University of East Anglia, Norwich, UK; [5]Cambridge Transplant Centre, Addenbrooke's Hospital, Cambridge, UK; [6]National Institute for Health Research (NIHR) Birmingham Biomedical Research Centre; [7]Institute of Immunology & Immunotherapy, University of Birmingham (UK); [8]Centre for Rare Diseases, Institute of Translational Medicine, University Hospitals Birmingham; [9]Toronto Centre for Liver Disease, University Health Network and University of Toronto, Toronto, Canada;[10]Translational Gastroenterology Unit, John Radcliffe Hospital, and Nuffield Department of Medicine, University of Oxford, UK; [11]Norwegian PSC Research Center, Department of Transplantation Medicine, Oslo University Hospital Rikshospitalet, Oslo, Norway; [12]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway; [13]Sheila Sherlock Liver Centre, Royal Free Hospital, London UK; [14]Liver Medicine and Transplantation Service, Freeman Hospital, Newcastle, UK; [15]Institute of Liver Studies, Kings College Hospital, London, UK; [16]Department of Hepatology, Leeds Teaching Hospital, Leeds, UK; [17]Scottish Liver Transplant Unit, Royal Infirmary of Edinburgh, Edinburgh, UK; [18]PSC Support, Oxfordshire, UK.

KEY WORDS: UK-PSC; prognostic factor, risk score; cholestasis; autoimmune liver disease

Correspondence

Dr Elizabeth C Goode

Department of Gastroenterology, Norfolk and Norwich University Hospital, Colney Lane, Norwich, UK, NR4 7UY.

+44 (0)1603 286286

ecg44@cam.ac.uk

Abbreviations

PSC; primary sclerosing cholangitis, ALP; alkaline phosphatase, HLA; human leukocyte antigen, GWAS; genome-wide association study, IBD; inflammatory bowel disease, UDCA; Ursodeoxycholic acid.

Financial Support

Financial support has been received by National Institute of Health Research (RD-TRC and Birmingham Biomedical Research Centre), Isaac Newton Trust, Addenbrooke's charitable trust, Norwegian PSC Research Center and PSC Support. GMH is supported by the Lily and Terry Horner Chair in Autoimmune Liver Disease Research, Toronto Centre for Liver Disease, Toronto.

Conflict of Interest

None of the authors declare any conflicts of interest.

Disclosures

GMH has been on advisory boards for PBC with GSK, Intercept and Novartis. He has been a study investigator for Gilead, Falk, NGM, BioTie, Shire, Intercept, GSK, Novartis and Cymabay. SMR has been on advisory boards for Intercept. PJT has been on advisory boards for Dr .Falk Pharma, Intercept and Tillots.

Author Contributions

Goode EC, Hirschfield GM and Rushbook SM were involved in study concept design, and oversight. Sandford RN, Alexander GJ, Chapman RW, and Walmsely M contributed to overall study inception and setup. Goode EC and Clark AB were involved in statistical analysis. Srivastava B, Mells GM, and Gelson WTH were involved in database design and management. Mells GM, Trivedi PJ, Williamson KD, Vesterhus MN, Castren E, Karlsen TH, Ji SG, Thorburn D, Hudson M, Heneghan M, Aldersley MA, Bathgate A, Hirschfield GM, Rushbrook SM and Anderson CA contributed data to the study. Goode EC, Clark AB, Hirschfield GM and Rushbrook SM wrote the manuscript and all its revisions; all authors approved submission.

Abstract

Background & Aims: We sought to identify factors predictive of liver transplantation or death in patients with Primary Sclerosing Cholangitis (PSC), and to develop and validate a contemporaneous risk score for use in a real-world clinical setting.

Methods: Analysing data from 1001 patients recruited to the UK-PSC research cohort, we evaluated clinical variables for their association with 2- and 10-year outcome through Cox-proportional hazards and C-statistic analyses. We generated risk scores for short- and long-

term outcome prediction, validating their use in two independent cohorts totalling 451 patients.

Results: 36% of the derivation cohort were transplanted or died over a cumulative follow-up of 7,904 years. Serum alkaline phosphatase ≥2.4×ULN at 1 year post diagnosis, was predictive of 10-year outcome (HR=3.05, C=0.63, median transplant-free survival 63 versus 108 months, p<0.0001), as was the presence of extra-hepatic biliary disease (HR=1.45, p=0.01). We developed two risk scoring systems based upon age, values of bilirubin, alkaline phosphatase, albumin, platelets, presence of extra-hepatic biliary disease and variceal haemorrhage, which predicted 2- and 10-year outcome with good discrimination (C=0.81 and 0.80 respectively). Both UK-PSC risk scores were well-validated in our external cohort, and out-performed the Mayo and APRI scores (C=0.75 and 0.63 respectively). Whilst heterozygosity for the previously validated HLA-DR*03:01 risk allele predicted increased risk of adverse outcome (HR=1.33, p=.001), its addition did not improve the predictive accuracy the UK-PSC risk scores.

Conclusions: Our analyses, based upon a detailed clinical evaluation of a large representative cohort of participants with PSC, furthers our understanding of clinical risk markers and reports the development and validation of a real-world scoring system to identify those patients most likely to die or require liver transplantation.

Introduction

Primary sclerosing cholangitis (PSC) is a chronic fibrosing cholestatic liver disease frequently associated with inflammatory bowel disease (IBD)(1). Disease progression culminates in end-stage liver disease and a high likelihood of death without liver transplantation(2). Patients with PSC have up to 15% lifetime risk of developing cholangiocarcinoma, which

parallels their risk of colorectal cancer(1,3-5). Insights into disease pathogenesis are limited but genetic studies highlight the importance of the adaptive immune system, with the strongest genetic association found within the HLA locus(6-8).

Clinical course is variable and efforts to individualise risk prediction are important for patients, clinicians and trials of experimental agents(9). Existing studies suggest that various clinical factors may predict the risk of adverse outcome. For example, elevated IgG4 concentration is reportedly associated (although not robustly validated) with an increased risk of progression to cirrhosis(10,11). Conversely, small-duct PSC confers an improved survival and lower risk of cholangiocarcinoma(12), as does a reduction in serum alkaline phosphatase (ALP) 1-2 years following diagnosis(13-15). Using cut-offs previously defined as stratifiers of risk in small bile duct disease primary biliary cholangitis (PBC), two studies have confirmed the independent prognostic value of ALP in PSC cohorts(13,14). However many studies evaluating risk prediction models in PSC have been limited by sample size, tertiary centre recruitment bias, failure to control for the interaction of variables with one another and lack of validation(16). With the exception of the revised Mayo Clinic model, prior prognostic models include histological staging(17-21). Whilst it is not unexpected that histology is a predictor of outcome (surrogates of liver fibrosis e.g. enhanced liver fibrosis score (ELF) and transient elastography perform equally well(22,23)), a simpler prognostic scoring is warranted. The revised Mayo Clinic model published in 2000, was designed to predict short-term survival within the proceeding 4 years and does not predict the need for transplantation(21). Updated scoring systems such as the Amsterdam-Oxford model are designed to predict PSC-related death and liver transplant but demonstrate only moderate predictive power (C=0.68), likely to be explained by limited study-cohort size(24).

Given our ability to capture the clinical characteristics of a large, clinically representative, cohort of patients with PSC through the United Kingdom NIHR Primary Sclerosing Cholangitis (UK-PSC) Rare Disease Translational Research Cohort, we sought to describe the clinical course of PSC and to identify clinical and genetic features early in disease course that are associated with increased risk of transplantation or death. In doing so, our subsequent internationally-validated findings provide clinically meaningful approaches to individualised risk prediction.

Methods

Study Design

Using data from patients recruited to the UK-PSC research cohort (www.uk-psc.com) we evaluated participants with PSC who were ≥18 years of age with PSC incident or prevalent between August 1, 2008 and March 31, 2015, including liver transplant recipients who had undergone transplantation for PSC at any point before March 31, 2015. Participants were recruited from throughout the UK across a research network of 155 collaborating National Health Service (NHS) Trusts, including nearly every hospital providing general or specialist hepatology services in the UK, excluding Northern Ireland.

Inclusion criteria were based upon accepted diagnostic criteria for PSC(25) and included the presence of cholestatic liver biochemistry tests with characteristic bile duct changes on either endoscopic retrograde cholangio-pancreatography (ERCP), magnetic resonance cholangio-pancreatography (MRCP) and/or liver histology. In order to address the challenges faced outside clinical trials of comparing MR images, the distinction between intrahepatic and extrahepatic biliary disease was determined by team review of local

radiographic reports of cholangiographic imaging, as opposed to single expert review. Involvement of first order bile ducts (right or left main hepatic duct) and/or common bile duct at cholangiography were classified as extrahepatic biliary disease, as opposed to their absence being classified as intrahepatic biliary disease. Exclusion criteria included congenital abnormalities of the biliary tree, previous biliary surgery likely to cause secondary sclerosing cholangitis, primary bile duct carcinoma, HIV cholangiopathy, PBC, positive anti-mitochondrial antibody, hepatic sarcoidosis and drug-induced liver injury.

The study was conducted in accordance with the guidelines of the Declaration of Helsinki and the principles of good clinical practice. All participants provided written informed consent. Multi-regional ethics committee (MREC) approval was granted by the Cambridgeshire 4 National Ethics committee (MREC Number 08/45/008) and by the research and development department of each collaborating hospital.

Data Capture

Data were collected onto pre-specified questionnaires through systematic review of case notes between March 31, 2013 and March 31, 2015. Data included patient demographics, diagnostic cholangiography/histology reports, serial biochemistry at diagnosis (t_0), 1-year following diagnosis (t_1) and 2-years following diagnosis (t_2), IBD status, concomitant autoimmune disease, use of ursodeoxycholic acid (UDCA), development of malignancy or liver decompensation, and progression to transplantation or death. 74% of all data was collected by the lead clinician researcher (EG) during site visits to each hospital, and completed questionnaires reviewed by a second clinician researcher for accuracy and missing data. The remaining 26% of data was collected by the responsible clinician or

research nurse at each hospital site, and reviewed by the lead clinician researcher (EG) upon return. Where patients were under the care of more than one hospital (e.g. a transplant centre and a general hospital), the questionnaire was sent sequentially to each hospital to ensure complete data capture. Missing or inaccurate data were systematically queried with the clinician who had completed the questionnaire to ensure complete and accurate data capture. Data that passed quality control were uploaded into a bespoke secure database.

Study Entry and Outcome

We calculated the time from PSC diagnosis to outcome event. t_0 was defined as the date of first cholangiographic imaging or liver biopsy demonstrating PSC. The first primary endpoint was liver transplantation, chosen as an important hard outcome for which a definitive time point is easily available. In the context of liver disease, it can be difficult to accurately define deaths solely related to liver disease, therefore the second primary endpoint of all-cause mortality was chosen as the most encompassing term that would include all liver related deaths. Participants who did not reach an endpoint were censored at the date of their most recent blood tests or follow-up. To ensure capture of all outcome events, we utilised the fact that every UK patient has a unique NHS number which ensures that clinical coding is linked across primary, secondary and tertiary care. This practice, in place throughout the 40-year study follow-up duration, ensures that the risk of missed events was minimal, and did not bias the analysis.

Explanatory Variables

We considered variables for their association with outcome and inclusion in the risk score, based upon clinical relevance or pre-existing evidence. To account for variability in the measurement of laboratory investigations, alkaline phosphatase (ALP) was taken as a ratio of the upper limit of normal range (ULN) for the reporting laboratory. Other laboratory measures used the following standard units of measurement; haemoglobin (Hb) g/l, platelet count (Plts) $\times 10^9$/l, albumin (Alb) g/l, bilirubin (Bili) µmol/l.

Clinical data analysis

We calculated and reported descriptive statistics as numbers or percentages. Variables with >40% missing data were excluded from further analysis. For this reason, the following variables were omitted from the analysis; INR, AST, IgG subclasses and date of first hepatic decompensation (ascites, hepatic encephalopathy, jaundice). Time-to-event analysis was conducted using Cox's proportional hazards model, ensuring at least 10 events per risk factor included in the model(26). To facilitate accurate risk prediction, events were truncated at 10 years of follow-up. To ensure sufficient variation within the dataset, categorical variables were only considered if the categories had >5% of the cohort in each category. Variables present in <5% of the cohort and thus excluded from the analysis were; smoking status and variceal haemorrhage at t_0.

We performed unadjusted/univariate analysis on the raw dataset to demonstrate associations between risk factors and outcome. To account for missing data, we performed multivariable imputation using iteratively chained equations, combined the results of ten imputed data sets using Rubin's equations, and estimated the adjusted/multivariable model

using this imputed dataset. We selected variables for the final risk score using backward elimination, with removal of risk factors not significant at the 10% level(27). Continuous variables were assessed for non-linear association using cubic splines. Variables demonstrating a linear association were included in a standard continuous fashion. Variables demonstrating a non-linear association were categorised using cubic splines and clinical judgement to allow for ease of interpretation.

Alkaline Phosphatase

We analysed the association between ALP at t_1 and t_2 with outcome, to determine the optimal threshold for predicting 10-year hazard of outcome. ALP was divided into categorical variables from </≥0.5 to 4×ULN, with increments of 0.1. We plotted each ALP cut-off against the hazard of reaching an endpoint. The optimal threshold for ALP was determined using Harrell's C statistic**.**

Derivation of the UK-PSC Risk Scores

We derived three separate risk scores, to determine the model with the best discrimination. The first was a score using t_0 data to predict 10-year risk of outcome, the second a short-term risk score using t_0 data to predict two-year risk of outcome (RS_{ST}), and the third, a long-term risk score using t_0 and t_2 data to predict 10-year risk of outcome (RS_{LT}). The RS_{LT} included only those patients not reaching a primary endpoint within 2 years of diagnosis. The discrimination of each score was compared using Harrell's C statistic. Calibration curves for RS_{ST} and RS_{LT} were generated by creating deciles of data and comparing the model's predicted with the observed rates in the cohort, estimated by the Kaplan-Meier curve.

Validation of the UK-PSC risk scores

We used data from two external PSC patient cohorts, not included in the original analysis, to validate the UK-PSC Risk Scores; the first a national validation cohort, n=352, from two UK hospitals (Transplant centre Queen Elizabeth Hospital Birmingham, and non-transplant centre John Radcliffe Hospital, Oxford(24)) and the second, an international validation cohort from Norway (n=99). Methods of validation cohort data collection were identical to the derivation cohort; retrospective data from individual auditing of electronic and paper case notes by clinician researchers followed by quality control. Validation of the scoring system was performed by fitting a Cox-model to the validation cohort using the scoring system derived from the derivation cohort(28). Further visual validation was performed by displaying Kaplan-Meier survival curves for risk groups in both cohorts(28). Risk groups were defined by dividing the derivation cohort into four equal sized groups with increasing RS_{LT}, and the validation cohort divided into four groups according to the same RS_{LT} categories.

Comparison of the UK-PSC score with existing scores

We analysed the predictive ability of the modified Mayo risk score and AST:platelet ratio index (APRI) scores in both derivation and validation cohorts, comparing them to the UK-PSC Risk Scores using Harrell's C statistic. Both the Mayo and APRI score algorithms include aspartate aminotransferase (AST). However, in most UK hospitals, AST is not measured as part of standard liver biochemical tests. AST was therefore not available for calculation of the Mayo risk or APRI scores. Other studies have demonstrated some equivalence of AST and ALT(29). Using a subset of patients for which both AST and ALT data were available at

the same time points, we demonstrated the correlation and concordance between the two variables. We then used ALT in place of AST in the calculation of the Mayo risk score and APRI scores.

Genetic data analysis

Prior genotyping was conducted using the Illumina Immunochip(7), a targeted genotyping array with dense marker coverage across 186 known disease loci from 12 immune-mediated diseases. We considered the following HLA risk alleles, known to be associated with PSC disease risk from GWAS studies(6, 7); *HLA-B*08:01*, and *HLA-DRB1*03:01, 04:01, 07:01, 13:01* and *15:01*. The association between HLA risk alleles and outcome were analysed using a test for trend across 0, 1 and 2 copies of each allele. We also tested for association between significant risk alleles and important clinical variables. After application of a Bonferroni correction, our threshold for statistical significance was $p<0.008$.

All analyses were performed using Stata software (version 14.0/SE; StataCorp LP, College Station, TX). This study was conducted and reported in accordance with TRIPOD statement for transparent reporting of a multivariate prediction model for individual prognosis or diagnosis(30).

Results

Cohort characteristics

1749 patients were recruited to the UK-PSC cohort; 1252 questionnaires distributed and 1131 returned. 130 were excluded following quality control (figure 1), leaving 1001 patients for analysis, recruited from 108 hospitals, including 7 transplant centres. 57% were

recruited from non-transplant centres. The cohort (table 1) included 64% male sex, diagnosed at a median age of 46.8 years with median follow-up of 14.8 years (range 0.2-40.4), censored at time of transplant. 44% had extrahepatic biliary disease and 72.5% had concomitant IBD, most commonly ulcerative colitis (80.4%) and 14.3% of the cohort had another autoimmune disease. UDCA was prescribed for 58% of the cohort within the first 2 years following diagnosis.

35.7% of patients reached a primary endpoint over a cumulative follow-up period of 7,904 years. 27.8% underwent liver transplantation at a median age of 47.0 years. 7.9% died without a transplant; 47.8% of all deaths were PSC-related. The overall proportion of the cohort who were event-free at 2, 5 and 10 years was 92%, 82% and 64% respectively. 39% of men reached an outcome, compared to 29% of females (χ^2 =10.07, p=0.002). 43% of those with extrahepatic biliary disease reached an outcome compared to 23% of those without (χ^2 =40.6, p=0.00). Patients with extrahepatic biliary disease had a reduced median transplant-free survival compared to those without extrahepatic biliary disease (11.7 versus 23 years). UDCA use in the first 2 years following diagnosis was not associated with outcome. 11% of patients developed a gastrointestinal cancer, most commonly colorectal (5.4%), followed by cholangiocarcinoma (3.3%).

Serum Alkaline Phosphatase is associated with PSC outcome

ALP data at t_1 and t_2 was available for 72% and 70% of the cohort respectively. At both timepoints, elevated ALP was associated with an increased 10-year hazard of reaching an outcome (p=<0.001) (figure 2a and 2b). There was a log-linear association between serum ALP and outcome, however for ease of interpretation we chose to categorise ALP using

cubic splines (supplementary figure 1). At t_1 the optimal threshold for predicting 10-year outcome was ALP≥2.4×ULN (HR=3.05, C=0.63) (figure 2c); where median transplant-free survival was 63 versus 108 months for those with ALP<2.4×ULN (p<0.0001 (log-rank test)) (figure 2e). At t_2, the optimal threshold for predicting 10-year outcome was ALP≥2.2×ULN (HR=3.05, C=0.66) (figure 2d), where median survival was 44 versus >96 months for those with a t_2 ALP<2.2×ULN (p<0.0001 (log-rank test)) (figure 2f).

Disease distribution is associated with outcome in PSC

Cholangiographic data at t_0 were available in 87.2% of the cohort. Presence of extrahepatic biliary disease was associated with adverse outcome (HR=1.45 (CI 1.09,1.92), p=0.010). Patients without extrahepatic biliary disease had improved 10-year event-free survival, although >50% of both groups were event free at 10-years and thus median survival was not reached (supplementary fig 2).

Derivation of a UK-PSC Risk Score

Our first UK-PSC risk score used factors available at t_0 to predict 10-year risk of outcome. Following multivariable analysis, 7 factors were included in the score; age at t_0, bilirubin, ALP, albumin, haemoglobin, platelets and presence of extrahepatic biliary disease a t_0 (supplementary table 1) (C=0.78, shrinkage=0.94). Our cohort demonstrated a high event rate (8%) within the first 2 years of diagnosis. Therefore, to determine if variables predicting short- and long-term risk differed, we derived a short-term risk score (RS_{ST}), using variables at t_0 to predict risk of outcome within 2 years following diagnosis; and a long-term risk score (RS_{LT}), using variables from t_2 to predict 10-year risk of outcome (see Box). Mean serum ALT, platelet count and ALP ratio were all significantly reduced from t_0 to t_2

(supplementary table 2). Both RS_{ST} and RS_{LT} demonstrated improved predictive ability over the original model.

Short-term UK-PSC Risk Score (RS_{ST}): Four variables at t_0 were associated with 2-year outcome; bilirubin, albumin, haemoglobin and platelet count (table 2). Based upon these coefficients, a prognostic model was developed to predict risk of death or liver transplantation by year 2 (C =0.81, shrinkage=0.92) (Box 1).

Long-term UK-PSC Risk Score (RS_{LT}): Seven variables at t_2 were association with 10-year risk of outcome; age at diagnosis, bilirubin at t_2, ALP at t_2, albumin at t_2, platelets at t_2, presence of extrahepatic biliary disease at t_0 and variceal haemorrhage by t_2 (C=0.80, shrinkage=0.96) (table 3 and Box 1). Calibration of RS_{ST} and RS_{LT} using predicted versus observed survival rates estimated by Kaplan-Meier curve demonstrated good correlation. The scores are available at http://www.uk-psc.com/riskscores.

To define low- and high-risk disease groups according to RS_{LT}, we divided the cohort into 4 four equal quartiles. Event-free survival, plotted on a Kaplan-Meir survival curve (figure 3a) demonstrated observed event rates of 6.0%, 8.4%, 19.1% and 55.8% in the four respective risk groups. Curves were generally well separated, although the model was less able to distinguish between the two lowest risk groups. RS_{LT} defining the four risk groups is shown in supplementary table 3.

Validation of the UK-PSC scores

We analysed the predictive ability of both risk scores in a separate national and international patient cohort. In the respective national and international validation cohorts (table 1), 62.4% and 75.7% of the cohort were male, diagnosed at a median age of 47 and 39 years, with 71% and 86% diagnosed with concomitant IBD. The most notable differences between the derivation and validation cohorts were the shorter median follow-up (6 and 8 years in the national and international cohorts, versus 14.8 years), higher death rate (25.3% and 21.2% versus 7.9%) and lower transplant rate (13.9% and 11.1% versus 27.8%).

Both the RS_{ST} and RS_{LT} were associated with outcome in the national validation cohort ($p<0.001$). The slope of the Cox model for the RS_{ST} in the validation cohort was 1.09 which is not significantly different from 1, indicating the discrimination was preserved. The slope for the RS_{LT} was 1.36 ($p=0.0071$) which is significantly different from 1, suggesting that the score is more predictive of outcome in the validation than the derivation cohort. In the international validation cohort, the lack of events within the first two years meant only the RS_{LT} could be validated. The RS_{LT} was associated with outcome in the international validation cohort ($p<0.001$); the slope was 1.60 and not significantly different from 1 ($p=0.014$), indicating preserved discrimination, although this was based upon only 37 individuals.

Further visual validation of the RS_{LT} was performed by comparing Kaplan-Meier survival curves for the validation cohort according to the same four previously defined risk groups as the derivation cohort (figures 3b and 3a respectively). Event rates were similar to the derivation cohort at 2.9%, 10.4%, 20.0% and 47.9% (supplementary table 3). Both set of

four curves were quite well separated, confirming that the model had discrimination in both cohorts, however the model was less able to distinguish between the two intermediate risk groups in the validation cohort.

Comparison of UK-PSC score with existing scores

We compared the predictive accuracy of the Mayo and APRI scores to the RS_{ST} and RS_{LT} in the imputed derivation dataset. Based upon a subset of 170 patients from the validation cohort, for which both AST and ALT measurement were available for t_0 and t_2, there was strong correlation, ($r = 0.94$, $p<0.0001$) and strong concordance ($c = 0.92$, $p<0.0001$) between the two variables. ALT was therefore used in place of AST for calculation of the Mayo and APRI scores. In predicting 2-year outcome, the RS_{ST} out-performed the APRI and Mayo scores with C statistics of 0.81, 0.63 and 0.75 respectively. In predicting 10-year outcome the RS_{LT} demonstrated an incremental improvement over the APRI and Mayo scores with C statistics of was 0.80, 0.59 and 0.79 respectively.

We then compared the predictive accuracy of the Mayo and MELD scores to the RS_{ST} and RS_{LT} in the validation dataset. In predicting 2-year outcome, the RS_{ST} out-performed the Mayo and MELD scores with C statistics of 0.81, 0.73 and 0.78 respectively. In predicting 10-year outcome the RS_{LT} demonstrated a markedly improved predictive accuracy compared with Mayo and MELD with C statistics of 0.85, 0.69 and 0.70 respectively.

HLA risk alleles are associated with outcome

HLA genotype was available for 635 patients. 27% and 9% of the cohort were heterozygous and homozygous for the *HLA-DR*03:01* risk allele respectively. Presence of this allele was associated with outcome in a dose-dependent manner (HR=1.33, CI 1.13, 1.58, p=0.001) (supplementary table 4). After testing for association between *HLA-DR*03:01* and clinical characteristics at diagnosis, we found *HLA-DR*03:01* risk alleles to be inversely correlated with mean age at diagnosis (no copies 47.6 years, heterozygote 46.6 years, homozygote 40.8 years, p=0.007)(supplementary table 5). Addition of the *HLA-DR*03:01* risk allele to the risk score did not improve the discrimination of the model. No association was observed with *HLA-B*08, HLA-DR*04:01, 07:01, 13:01* or *15:01.*

Discussion

Using a large cohort of 1001 patients from across the entire UK, including 57% recruited from non-transplant centres and thus representing the full spectrum of PSC disease severity, we provide important, externally validated clinical and genetic modelling, based upon readily available clinical factors for prediction of short and long-term outcome. Based upon presence of extrahepatic biliary disease at t_0, age, bilirubin, ALP, albumin and platelets at t_2 and variceal haemorrhage by t_2, we present a scoring system both of value in individual risk evaluation, as well as a potential mechanism to stratify recruitment to clinical trials.

Our study confirms the importance of ALP as a prognostic indicator, both individually and as part of our RS_{LT}. We demonstrate ALP<2.4×ULN and <2.2×ULN at 1 and 2 years following diagnosis, is associated with improved transplant–free survival. Understanding the

behaviour of ALP as a biomarker in PSC is of interest, and parallels interest in PBC. In PBC, dichotomous risk scores have C statistics of around 0.6, with the dynamic scores reporting C statistics of 0.8 and above(31, 32). Whilst in PSC, serum ALP was not associated with short-term outcome, the prognostic importance as longer-term predictor of clinical events is highlighted by its inclusion in our long-term risk score. This may be explained by fluctuations in ALP in the early stages of diagnosis, which limit its prognostic value, and the rationale that short-term risk is driven by factors that measure cholestasis and portal hypertension. Thus, when considering ALP in isolation, we chose to consider ALP at t_1 and t_2 rather than t_0, and used this to predict long-term rather than short-term risk. In addition, our study demonstrated the poor prognostic impact of extrahepatic biliary disease, highlighting the importance of further study into cholangiographic monitoring in PSC. Whilst simple cholangiographic imaging used at diagnosis does carry meaningful prognostic data, improving cholangiographic evaluation remains important.

We observed a high event rate (8%) within the first two years following diagnosis, suggesting there is a patient cohort who present late in disease course, or who experience rapidly progressive disease. Recognising this, we developed separate risk scores for short- (RS_{ST}) and long-term (RS_{LT}) prediction, the key differences between which are that the former includes only laboratory parameters (bilirubin, albumin, haemoglobin and platelets), suggesting that intrinsic liver function is most important in predicting immediate outcome. Conversely the RS_{LT} includes laboratory factors (bilirubin, albumin, platelets, ALP) in addition to variceal bleeding and cholangiographic disease distribution. By using a dichotomous approach to risk stratification, we improved predictive utility from C=0.78 with our original score, to C=0.80 and 0.81 for short-term prediction and C=0.80 and 0.85 for long-term

prediction, in the derivation and validation cohorts respectively. In practice, this would allow clinicians to recalculate risk at 2 years following diagnosis for greater prognostic accuracy.

There must remain a risk of better performance of our model, due to data fitting, a risk we acknowledge. We tried to address this by comparing our risk scores existing scores, including the Mayo score. Whilst we reconfirm the Mayo risk score's prognostic value, it was out-performed by our RS_{ST} and RS_{LT}, which also confer several other advantages. Whilst the Mayo score is based upon the parameters age, bilirubin, AST, variceal bleeding and albumin, the UK-PSC risk scores consider more aspects of disease progression including age, ALP, albumin, platelets, extra-hepatic biliary disease and variceal haemorrhage. The Mayo risk score predicts only 4-year risk of all-cause mortality and does not provide a strong long-term predictor of outcome. It performs best in patients with end-stage liver disease and does not consider the important outcome of liver transplantation. In comparison, the dichotomous UK-PSC risk scores predicts short- (2-year) and long-term (10-year) outcomes, ensuring that predictive ability is as good for those patients presenting with early, as well as late stage liver disease and includes the important outcome of liver transplantation in addition to all-cause mortality.

We found one previously validated HLA risk allele to be an important in predictor of disease outcome: *HLA-DR*03:01*, demonstrated a gene-dose effect; but notably the addition of *HLA-DR*03:01* did not improve the predictive ability of our prognostic score. When considering effect size, *HLA-DR*03:01* had an adjusted hazard ratio of 1.33 (CI 1.13,1.58, p=0.001) for outcome, more comparable to that of extrahepatic biliary disease or ALP, in

comparison with the strongest associated clinical variable, bilirubin which had an adjusted odds ratio of 3.96 (CI 2.4,6.9, p<0.001).

A strength of our cohort is the representative nature of the participants, notably identifying low-risk as well as high-risk patients. In our study, we can confidently define a 'low risk' disease group according to RS_{LT} (patients with a RS_{LT} of <-2.02 had <10% chance of an event by 10 years follow up) and a 'high risk' group (patients with RSLT score -0.81 < RSLT < 2.74 had ~ 50% chance of an event by 10 years). Both UK-PSC Risk scores (http://www.uk-psc.com/riskscores) were well validated in a separate patient cohort. The major difference between derivation and validation cohorts was a lower death and higher transplant rate in the former. There are some biases in our derivation cohort, reflective of ascertainment processes. Recruitment to the UK-PSC derivation cohort was retrospective through prevalent case ascertainment. Recruitment was therefore inherently biased towards those patients with or without transplant, who survived to 2008 to be recruited to the study, compared to those patients who died before 2008. A further function of this, is the low prevalence of cholangiocarcinoma (3.3%) in our cohort. Retrospective cohort recruitment is not well suited to capturing data on outcome markers associated with very poor survival; nearly 50% of all PSC-associated cholangiocarcinomas manifest within 2 years of PSC diagnosis(2). Despite these limitations, the UK-PSC risk scores were nationally and internationally well-validated in two external cohorts, lending weight to their importance as robust prognostic models.

Retrospective data collection also carries the inherent drawback of incomplete data collection. Not surprisingly, rates of missing data were higher for patients diagnosed many years previously. Given that it was only related to the year of diagnosis, we considered these data to be missing at random, and used imputation to improve the validity of the results. Whilst a date of diagnosis before 1990 was associated with an improved transplant-free survival, removal of these patients from the analysis, did not alter the strength of the reported associations with either short- or long-term outcome, and thus they were retained for the purpose of statistical power.

In our study, we did not observe any difference in outcome according to sex or subtype of IBD. The evidence for female sex and Crohn's disease conferring a favourable outcome in patients with PSC has only been robustly supported by evidence from one large study including more than 7000 patients with PSC(2). It is therefore likely that with a total sample size of 1452, our study was underpowered to detect any such an effect. Further studies of even larger cohorts, adjusted for multiple factors, are needed to confidently validate this finding. In particular, this necessitates careful consistent classification of PSC-associated IBD.

We chose the endpoints of all-cause mortality and liver transplantation rather than hepatic decompensation (e.g. ascites and hepatic encephalopathy), because they provide a definitive and easily identifiable event and time of event, necessary for a retrospective observational study. In comparison, hepatic decompensation can remain undiagnosed for several months and the precise date of diagnosis remain subjective; such issues are less

relevant in well-designed clinical trial/prospective cohort settings where such endpoints are clearly meaningful and can be collected robustly. Arguably there are challenges with our chosen endpoints due to possible variation in clinical practice and outcomes over time. Indeed, it must always be acknowledged that changes in disease course over the time period of a study needed to evaluate outcomes, is a potential confounding factor. Evaluating such changes can be hard; contemporaneous reference literature may, for example, not reflect clinical changes to a disease manifestation, but reporting and investigative practices. Despite PSC being infrequent overall, and the difficulty in recruiting large cohorts for the development of well-powered evaluations, we have collated a unique dataset capturing ~15% of the total UK PSC population, with substantial power to evaluate all-cause mortality and transplantation. Whilst accepting limitations inherent to our approach there is no evidence to date, that PSC outcomes have varied over the time course of our study, simply on the basis of era, and recent data reporting liver transplant practice in the UK additively supports this(33). Our approach is importantly, of significance to clinicians as the risk score analyses best reflect collective real-world clinical practice, with a focus on a spectrum of patients reflective of the disease, and a study design that through large cohort size, accommodates the weaknesses introduced by non-trial based uniform evaluation and data capture.

In conclusion, our analyses based on a detailed clinical evaluation of a large representative cohort of participants with PSC has furthered our understanding of clinical risk markers for predicting outcome in patients with PSC.

Acknowledgements

The authors gratefully acknowledge the work of UK-PSC Consortium members (see Supplementary file). We thank all participants who granted access to their medical records, enabling us to conduct this study. The UK-PSC study is a portfolio study of the National Institute for Health Research (NIHR) Comprehensive Research Network. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. We thank the ongoing support of UK patient support group PSC Support.

References

1. Bergquist A, Ekbom A, Olsson R, Kornfeldt D, Loof L, Danielsson A, Hultcrantz R, et al. Hepatic and extrahepatic malignancies in primary sclerosing cholangitis. J Hepatol 2002;36:321-327.

2. **Weismuller TJ, Trivedi PJ,** Bergquist A, Imam M, Lenzen H, Ponsioen CY, Holm K, et al. Patient Age, Sex, and Inflammatory Bowel Disease Phenotype Associate With Course of Primary Sclerosing Cholangitis. Gastroenterology 2017; 152(8):1975-1984.

3. Sano H, Nakazawa T, Ando T, Hayashi K, Naitoh I, Okumura F, Miyabe K, et al. Clinical characteristics of inflammatory bowel disease associated with primary sclerosing cholangitis. J Hepatobiliary Pancreat Sci 2011;18:154-161.

4. Claessen MM, Vleggaar FP, Tytgat KM, Siersema PD, van Buuren HR. High lifetime risk of cancer in primary sclerosing cholangitis. J Hepatol 2009;50:158-164.

5. Razumilava N, Gores GJ, Lindor KD. Cancer surveillance in patients with primary sclerosing cholangitis. Hepatology 2011;54:1842-1852.

6. Karlsen TH, Franke A, Melum E, Kaser A, Hov JR, Balschun T, Lie BA, et al. Genome-wide association analysis in primary sclerosing cholangitis. Gastroenterology 2010;138:1102-1111.

7. Liu JZ, Hov JR, Folseraas T, Ellinghaus E, Rushbrook SM, Doncheva NT, Andreassen OA, et al. Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. Nat Genet 2013;45:670-675.

8. Ji SG, Juran BD, Mucha S, Folseraas T, Jostins L, Melum E, Kumasaka N, et al. Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. Nat Genet 2017;49:269-273.

9. Trivedi PJ, Corpechot C, Pares A, Hirschfield GM. Risk stratification in autoimmune cholestatic liver diseases: Opportunities for clinicians and trialists. Hepatology 2016;63:644-659.

10. Broome U, Lofberg R, Veress B, Eriksson LS. Primary sclerosing cholangitis and ulcerative colitis: evidence for increased neoplastic potential. Hepatology 1995;22:1404-1408.

11. Bjornsson E, Chari S, Silveira M, Gossard A, Takahashi N, Smyrk T, Lindor K. Primary sclerosing cholangitis associated with elevated immunoglobulin G4: clinical characteristics and response to therapy. Am J Ther 2011;18:198-205.

12. Bjornsson E, Boberg KM, Cullen S, Fleming K, Clausen OP, Fausa O, Schrumpf E, et al. Patients with small duct primary sclerosing cholangitis have a favourable long term prognosis. Gut 2002;51:731-735.

13. Al Mamari S, Djordjevic J, Halliday JS, Chapman RW. Improvement of serum alkaline phosphatase to <1.5 upper limit of normal predicts better outcome and reduced risk of cholangiocarcinoma in primary sclerosing cholangitis. J Hepatol 2013;58:329-334.

14. Lindstrom L, Hultcrantz R, Boberg KM, Friis-Liby I, Bergquist A. Association between reduced levels of alkaline phosphatase and survival times of patients with primary sclerosing cholangitis. Clin Gastroenterol Hepatol 2013;11:841-846.

15. Stanich PP, Bjornsson E, Gossard AA, Enders F, Jorgensen R, Lindor KD. Alkaline phosphatase normalization is associated with better prognosis in primary sclerosing cholangitis. Dig Liver Dis 2011;43:309-313.

16. Boonstra K, Weersma RK, van Erpecum KJ, Rauws EA, Spanier BW, Poen AC, van Nieuwkerk KM, et al. Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. Hepatology 2013;58:2045-2055.

17. Broome U, Olsson R, Loof L, Bodemar G, Hultcrantz R, Danielsson A, Prytz H, et al. Natural history and prognostic factors in 305 Swedish patients with primary sclerosing cholangitis. Gut 1996;38:610-615.

18. Wiesner RH, Grambsch PM, Dickson ER, Ludwig J, MacCarty RL, Hunter EB, Fleming TR, et al. Primary sclerosing cholangitis: natural history, prognostic factors and survival analysis. Hepatology 1989;10:430-436.

19. Farrant JM, Hayllar KM, Wilkinson ML, Karani J, Portmann BC, Westaby D, Williams R. Natural history and prognostic variables in primary sclerosing cholangitis. Gastroenterology 1991;100:1710-1717.

20. Dickson ER, Murtaugh PA, Wiesner RH, Grambsch PM, Fleming TR, Ludwig J, LaRusso NF, et al. Primary sclerosing cholangitis: refinement and validation of survival models. Gastroenterology 1992;103:1893-1901.

21. Kim WR, Therneau TM, Wiesner RH, Poterucha JJ, Benson JT, Malinchoc M, LaRusso NF, et al. A revised natural history model for primary sclerosing cholangitis. Mayo Clin Proc 2000;75:688-694.

22. Corpechot C, Gaouar F, El Naggar A, Kemgang A, Wendum D, Poupon R, Carrat F, et al. Baseline values and changes in liver stiffness measured by transient elastography are associated with severity of fibrosis and outcomes of patients with primary sclerosing cholangitis. Gastroenterology 2014;146:970-979.

23. Vesterhus M, Hov JR, Holm A, Schrumpf E, Nygard S, Godang K, Andersen IM, et al. Enhanced liver fibrosis score predicts transplant-free survival in primary sclerosing cholangitis. Hepatology 2015;62:188-197.

24. de Vries EM, Wang J, Williamson KD, Leeflang MM, Boonstra K, Weersma RK, Beuers U, et al. A novel prognostic model for transplant-free survival in primary sclerosing cholangitis. Gut 2017. doi: 10.1136/gutjnl-2016-313681

25. Chapman R, Fevery J, Kalloo A, Nagorney DM, Boberg KM, Shneider B, Gores GJ, et al. Diagnosis and management of primary sclerosing cholangitis. Hepatology 2010;51:660-678.

26. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol 2007;165:710-718.

27. Mallett S, Royston P, Dutton S, Waters R, Altman DG. Reporting methods in studies developing prognostic models in cancer: a review. BMC Med 2010;8:20.

28. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013;13:33.

29. Carbone M, Sharp SJ, Flack S, Paximadas D, Spiess K, Adgey C, Griffiths L, et al. The UK-PBC risk scores: Derivation and validation of a scoring system for long-term prediction of end-stage liver disease in primary biliary cirrhosis. Hepatology 2015.

30. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). Ann Intern Med 2015;162:735-736.

31. Lammers WJ, Kowdley KV, van Buuren HR. Predicting outcome in primary biliary cirrhosis. Ann Hepatol 2014;13:316-326.

32. Lammers WJ, van Buuren HR, Hirschfield GM, Janssen HL, Invernizzi P, Mason AL, Ponsioen CY, et al. Levels of alkaline phosphatase and bilirubin are surrogate end points of outcomes of patients with primary biliary cirrhosis: an international follow-up study. Gastroenterology 2014;147:1338-1349.

33. Webb GJ, Rana A, Hodson J, Akhtar MZ, Ferguson JW, Neuberger JM, Vierling JM. Twenty-year comparative analysis of patients with autoimmune liver diseases on transplant waitlists. Clin Gastroenterol Hepatol 2018;16:278-287.

Figure legends

Figure 1: Study flow diagram

Figure 2: Predictive value of Alkaline Phosphatase and outcome

a) Association between alkaline phosphatase (as ratio of ULN) at year 1 and hazard of reaching a clinical endpoint within 10 years, with 95% CI **b)** Association between alkaline phosphatase (as ratio of ULN) at year 2 and hazard of reaching a clinical endpoint within 10 years, with 95% CI **c)** Harrell's C- statistic per ALP cut-point at year 1 for 10-year hazard of outcome **d)** Harrell's C- statistic per ALP cut-point at year 2 for 10-year hazard of outcome **e)** Kaplan Meier survival curve for transplant-free survival in patients with ALP≤2.4×ULN (blue line) versus ALP>2.4×ULN (red line) at 1 year following diagnosis (0 = 12 months post diagnosis) **f)** Kaplan Meier survival curve for transplant-free survival in patients with ALP≤2.2×ULN (blue line) versus ALP>2.2×ULN (red line) at 2 years following diagnosis (0 = 24 months post diagnosis)

Figure 3: Kaplan-Meir survival curves for 4 risk groups

Risk group 1; RS_{LT} > -2.019879 (blue line), risk group 2; -1.463874 <RS_{LT} < -2.019879 (red line), risk group 3; -0.8146346 <RS_{LT} < -1.463874 (green line), risk group 4; 2.737384 <RS_{LT} < -0.8146346 (orange line). **a)** Derivation cohort Kaplan-Meir survival curves for 4 risk groups **b)** Validation cohort Kaplan-Meir survival curves for 4 risk groups

Table 1: Demographics of the UK-PSC derivation cohort (n=1001), national validation cohort (n=352) and international validation cohort (n=99)

	Variable	Derivation Cohort n=1001 (%)	Validation Cohorts National n=352 (%)	International n=99 (%)
Demographics	Male	63.8	62.4	75.7
	Mean age at diagnosis (yrs)	46.8	45.0	35.0
	Median age at transplant	47	47.0	39.0
	Median follow-up (yrs)	14.8	6.0	8.0
Disease Distribution	Extrahepatic biliary disease present	44.1	47.8	33.3
IBD	IBD	72.5	71.0	86.0
	Ulcerative Colitis	80.4	73.6	77.6
	Crohns Colitis	14.2	10.7	15.3
	Indeterminate Colitis	5.4	3.2	7.1
Autoimmune disease	Autoimmune disease	14.3	-	-
	Thyroid disease	6.9	-	-
	Rheumatoid arthritis	2.3	-	-
	Coeliac Disease	2.0	-	-
	Other	6.2	-	-
Smoking status	Never smoked	53.2	-	-
	Ex-smoker	26.5	-	-
	Current smoker	3.7	-	-
Events	Total events	35.7	39.2	32.3
	Transplants	27.8	13.9	11.1
	Deaths (all-cause)	7.9	25.3	21.2
Cancers	GI Cancer	10.7	-	-
	Colorectal Ca	5.4	-	-
	Cholangio Ca	3.3	-	-
	GB Cancer	1.3	-	-
	HCC	0.6	-	-
	Pancreatic Cancer	0.1	-	-
UDCA use	Taking UDCA at year 2	57.8	-	-
	Median dose (mg/kg)	11.4	-	-
	Range (mg/kg)	2.2-46.8	-	-

PSC; Primary Sclerosing Cholangitis, IBD; inflammatory bowel disease, UDCA; Ursodeoxycholic acid, GI; Gastrointestinal, GB; gallbladder, HCC; Hepatocellular carcinoma, Ca; cancer, UDCA; ursodeoxycholic acid

Table 2: Univariate analysis using un-imputed data and multivariate analysis using imputed data, of factors at **diagnosis** associated with **2-year risk** of transplantation or death

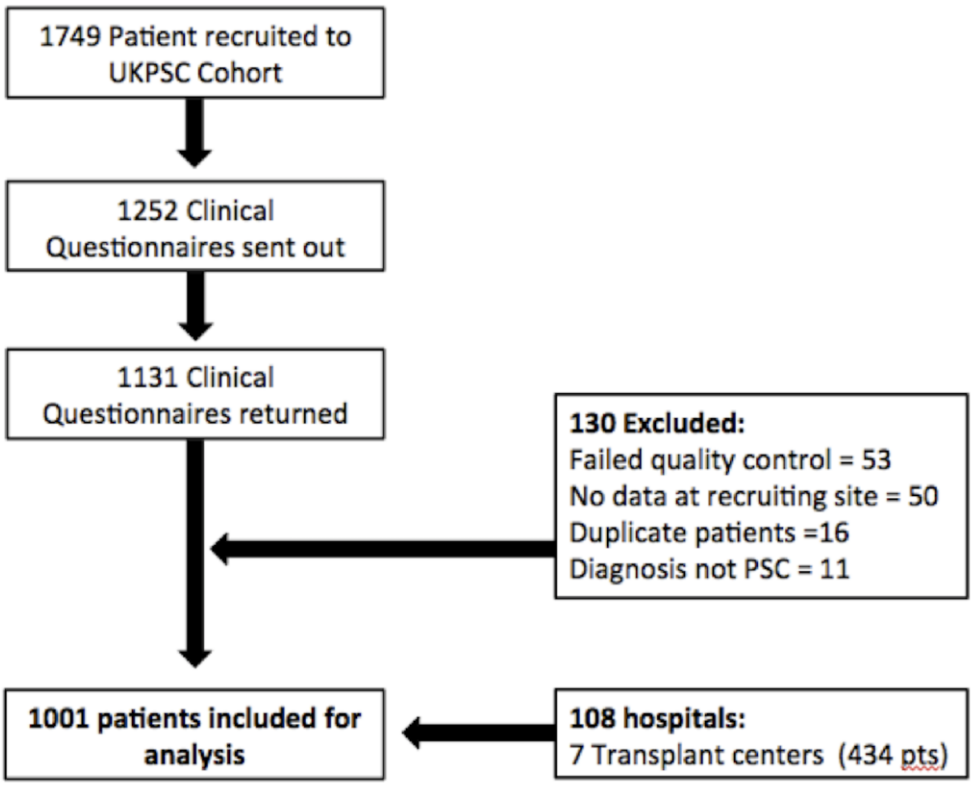
	Univariate Analysis		Multivariate Analysis	
Factor	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**
Female	0.88 (0.54,1.42)	0.596		
Age at diagnosis	1.01 (1.00,1.03)	0.126		
Extrahepatic biliary disease	1.30 (0.77,2.21)	0.332		
IBD presence	1.09 (0.49,2.44)	0.832		
UC	1.12 (0.67,1.89)	0.665		
CD	0.39 (0.12,1.31)	0.127		
Indeterminate colitis	1.38 (0.47,4.03)	0.560		
Autoimmune disease	0.90 (0.46,1.75)	0.757		
Smoker	1.22 (0.74,2.02)	0.426		
Bilirubin (µmol/l)				
35 -49	4.03 (1.36,11.98)	0.012	2.11 (0.74,5.96)	0.159
50+	14.12 (7.89,25.3)	<0.001	5.02 (2.76,9.13)	<0.001
ALP (ratio of ULN)				
1.5 - <2.5	1.25 (0.49,3.17)	0.634		
2.5+	2.64 (1.35,5.17)	0.005		
ALT (IU/l)*	1.02 (0.98,1.05)	0.331		
Albumin (g/l)	0.87 (0.84,0.90)	<0.001	0.94 (0.90,0.99)	0.011
Haemoglobin (g/l)**	0.98 (0.97,0.99)	<0.001	0.99 (0.97,1.00)	0.095
Platelets group ($\times 10^9$/l)				
150 - 199	0.23 (0.08,0.72)	0.011	0.62 (0.26,1.48)	0.283
200 - 399	0.22 (0.11,0.45)	<0.001	0.50 (0.25,0.98)	0.045
400+	0.32 (0.13,0.78)	0.012	0.38 (0.15,0.98)	0.046
Eosinophils ($\times 10^9$/l)	1.10 (0.89,1.36)	0.368		
Sodium (mmol/l)	0.89 (0.82,0.98)	0.015		
Creatinine >120 (µmol/l)	4.21 (1.66,10.68)	0.002		
IgG (g/l)*	1.08 (0.93,1.25)	0.313		

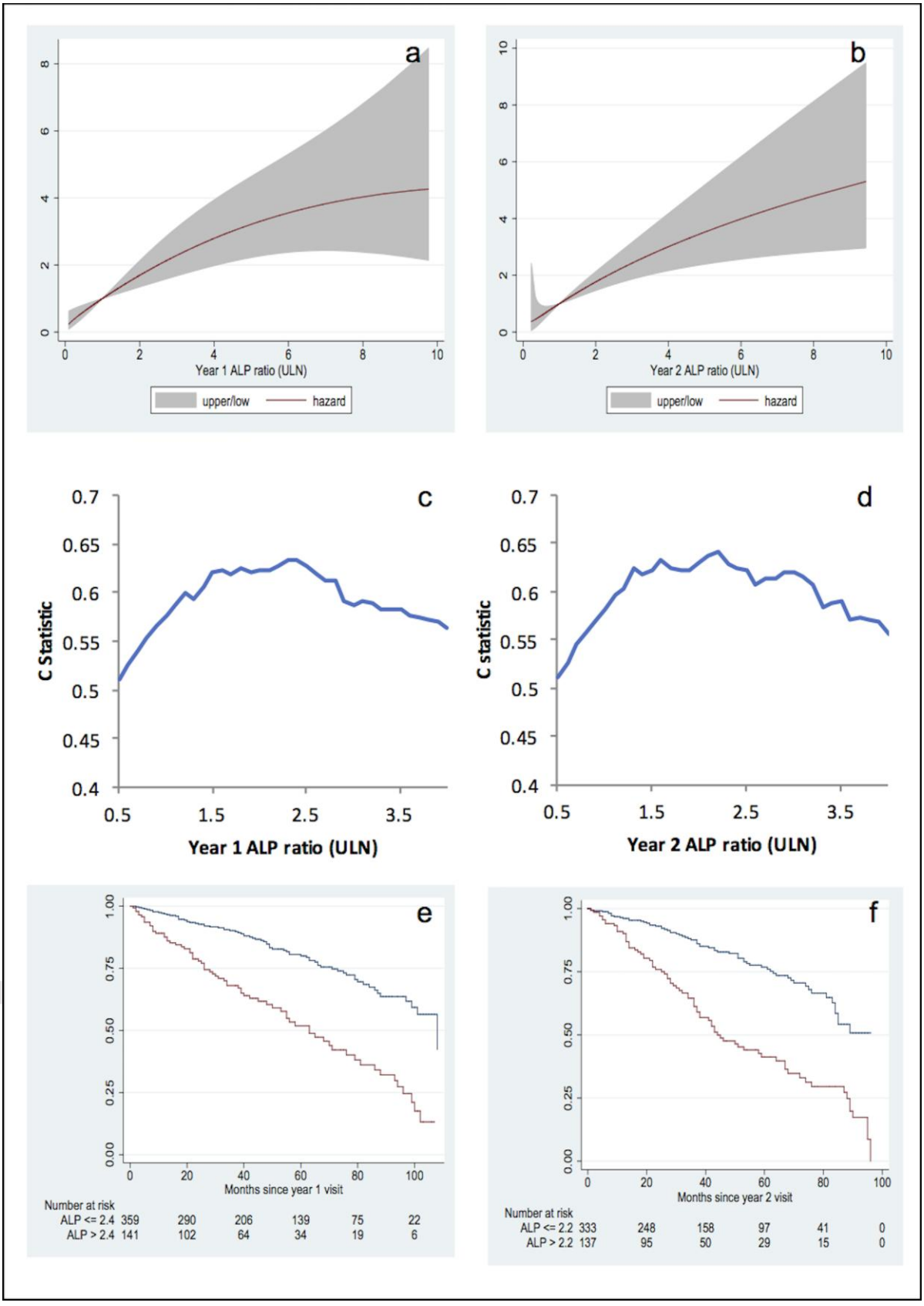
IBD; Inflammatory bowel disease, UC; Ulcerative colitis, CD; Crohns disease, ALP; Alkaline Phosphatase ALT; Alanine Transaminase, IgG; immunoglobulin G, *denotes hazard ratio for a 10 unit change, **denotes hazard ratio for a 1 unit change

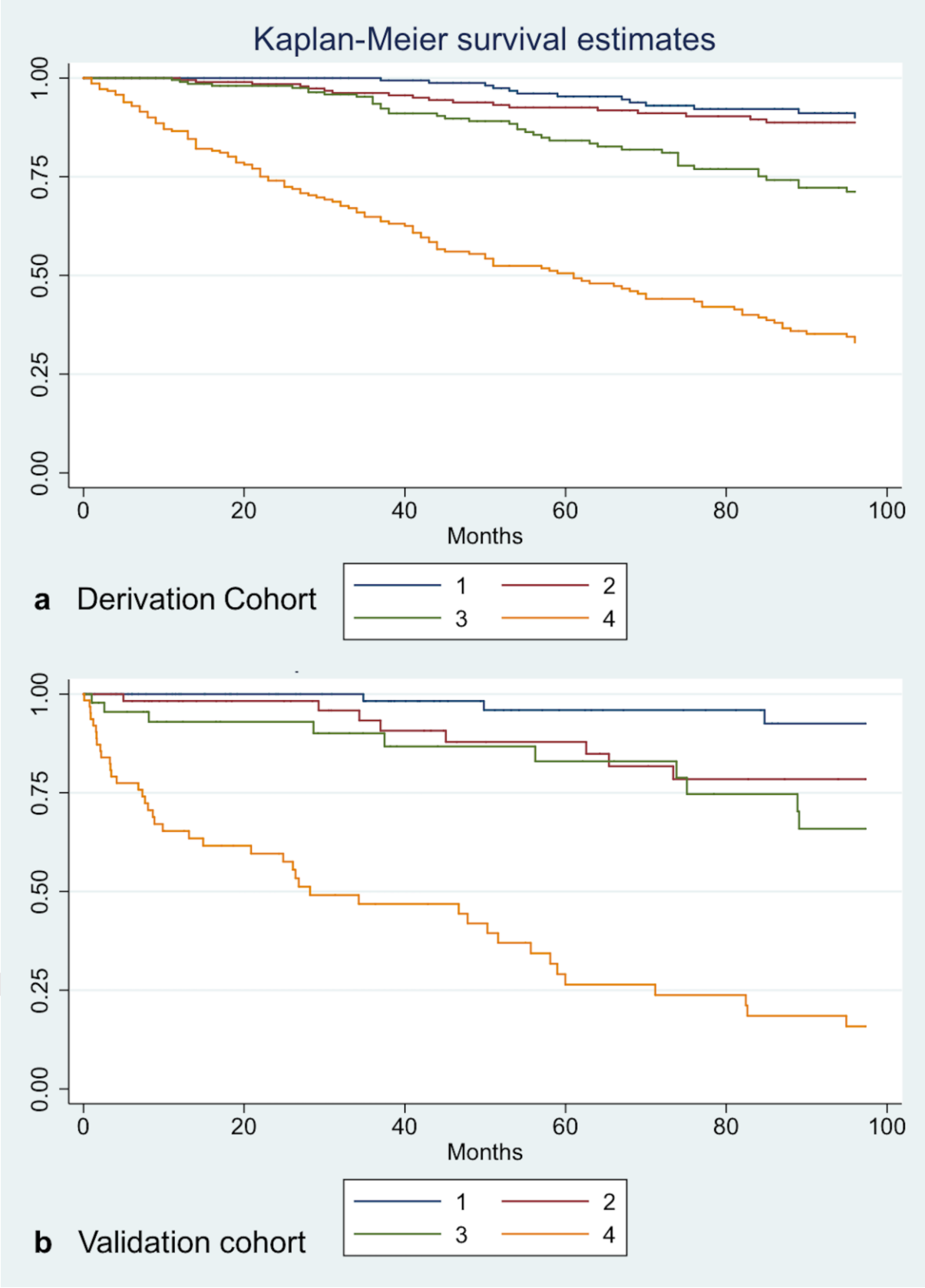
Table 3: Univariate analysis using un-imputed data and multivariate analysis using imputed data, of factors at **year 2** associated with **10-year** risk of transplantation or death

	Univariate Analysis		Multivariate Analysis	
Factor	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**
Female	0.81 (0.60,1.10)	0.181		
Age at diagnosis	1.01 (1.00,1.03)	0.005	1.03 (1.01,1.04)	<0.001
Extrahepatic biliary disease	1.95 (1.42,2.69)	<0.001	1.70 (1.15,2.48)	0.008
IBD	0.91 (0.59,1.38)	0.646		
UC	0.92 (0.70,1.22)	0.558		
CD	0.68 (0.41,1.11)	0.122		
Indeterminate colitis	1.28 (0.71,2.31)	0.416		
Autoimmune disease	1.27 (0.88,1.83)	0.200		
Smoker	0.96 (0.70,1.32)	0.790		
Bilirubin (µmol/l)				
35 -49	6.77 (3.87,11.85)	<0.001	3.31 (1.65,6.62)	0.001
50+	7.92 (5.62,11.18)	<0.001	3.96 (2.37,6.62)	<0.001
ALP (ratio of ULN)				
1.5 - 2.4	1.75 (0.98,3.15)	0.061	1.50 (1.09,2.30)	0.015
2.5+	1.40 (1.04,1.88)	0.025	1.57 (1.12,2.52)	0.011
ALT (IU/l)**	1.05 (1.03,1.08)	<0.001		
Albumin (g/l)	0.88 (0.85,0.90)	<0.001	0.93 (0.90,0.96)	<0.001
Haemoglobin (g/l)**	0.75 (0.69,0.81)	<0.001		
Platelets group ($\times 10^9$/l)				
150 - 199	0.35 (0.20,0.60)	<0.001	0.58 (0.31,1.10)	0.092
200 - 399	0.29 (0.20,0.43)	<0.001	0.60 (0.40,0.91)	0.016
400+	0.32 (0.17,0.60)	<0.001	0.46 (0.23,0.92)	0.028
Eosinophils ($\times 10^9$/l)	0.81 (0.52,1.29)	0.380		
Sodium (mmol/l)	0.90 (0.96,0.93)	<0.001		
Creatinine >120 (µmol/l)	0.66 (0.21,2.07)	0.474		
IgG (g/l)**	1.01 (0.92,1.12)	0.774		
UDCA use	0.96 (0.72,1.28)	0.795		
Variceal bleed by yr 2	5.97 (2.93,12.16)	<0.001	2.76 (1.14,6.66)	0.024

**Denotes hazard ratio for a 10 unit change







Box 1:

Short-term UK-PSC Risk Score (R_{ST}) = 0.745(Bili_t_0 Group 1 [0/1] + 1.613(Bili_t_0 Group 2 [0/1]) – 0.061(Alb_t_0 [g/l]) - 0.012(Hb_t_0 [g/l]) -0.476(Plts_t_0 Group 1 [0/1]) – 0.698(Plts_t_0 Group 2 [0/1]) -0.962(Plts_t_0 Group 3 [0/1]).

Long-term UK-PSC Risk Score (RS_{LT}) = 0.026(Age_t_0[yrs]) + 1.197(Bili_t_2 Group 1 [0/1]) + 1.38(Bili_t_2 Group 2 [0/1]) + 0.4(ALP_t_2 Group 1 [0/1]) + 0.45(ALP_t_2 Group 2 [0/1]) – 0.07(Alb_t_2[g/l]) –0.543(Plts_t_2 Group 1) – 0.503(Plts_t2 group 2) – 0.768(Plts_t_2 Group 3 [0/1]) + 0.524(disease type_t_0 [0/1]) + 1.014(variceal bleed_t_2 [0/1]).

***Bili_t_0/t_2 group 1**; 0= Bili_t_0<35 µmol/l or >50 µmol/l, 1= 35 to ≤ 50 µmol/l*
***Bili_t_0/t_2 group 2**; 0=Bili_t_0 < 50 µmol/l, 1=Bili_t_0 ≥50 µmol/l*
***Plts_t_0/t_2 group 1**; 0=Plts_t_0<150×10^9/l, or ≥200×10^9/l, 1= Plts_t_0 150 to <200×10^9/l*
***Plts t_0/t_2 group 2**; 0=<200 or ≥ 400×10^9/l, 1= 200 to <400×10^9/l*
***Plts_t_0/t_2 group 3**; 0=<400×10^9/l, 1= ≥ 400×10^9/l*
***ALP_t_2 group 1**; 0=ALP_t_2<1.5×ULN or ≥2.5×ULN, 1=1.5 to <2.5,*
***ALP_t_2 group 2**; 0= ALP_t_2<2.5×ULN, 1=≥2.5×ULN,*
***Disease type_t_0**; 0=no extrahepatic biliary disease, 1=presence of extrahepatic biliary disease **Variceal bleed_t_2**; 0= no bleed by t_2, 1=bleed by t_2.*

Predicted survival rate at time t = (Baseline survival at time t) ^ exp (RS_{ST} or RS_{LT}),

RS_{ST} baseline survival at time t; 1 year: 0.0096612, 2 years: 0.0001109
RS_{LT} baseline survival at time t; 1 year: 0.9218476, 2 years: 0.8227174, 5 years: 0.7070919, 8 years: 0.2771266.

Example
An individual aged 47 and with no evidence of extra-hepatic disease at diagnosis with the following biochemistry at t_0: Bili 37µmol/l, Alb 34 g/l, Hb 130 g/l and Plts 245×10^9/l, and the following biochemistry at t_2: Bili 24µmol/l, ALP 2×ULN, Alb 30 g/l, Plts 152×10^9/l and no variceal bleed by t_2, would score would score: RS_{ST}= (0.745 ×1) – (0.061 ×34)-(0.012 ×130)-(0.698×1)= -3.587. Predicted event free survival rate at 2 years = (0.0001109)^exp (-3.587)= 0.78= 78%. RS_{LT}=(0.026 × 47) + 0.403 + (- 0.07 × 30)– 0.543 = -1.018. Predicted event free survival rate at 5 years = 0.707 ^ exp (-1.047) = 0.885 =88.5% .

The UK-PSC risk scores are available at http://www.uk-psc.com/riskscores